

For continuous type measurements, our approach to assessing inter-examiner consistency uses a balanced paired assessment design in which a sample of participants are each measured by a pair of examiners. To increase power for detecting examiner inconsistency, enough participants are used so that each examiner pairs up with each other examiner at the same clinic center a fixed small number of times (usually one to three), depending on the number of examiners at the clinical center and the desired detectable examiner effect size. In this situation, the effect size is the variance of between-examiner differences divided by the within-examiner variance. We usually design the size of the study so there is good power for detecting when this ratio is one or more. For example, for a power of 80% to detect examiner differences on the same order of magnitude as within-examiner variation, requires 2 repeat measurements per pair of examiners when there are 5 examiners (10 possible examiner pairs), or 20 subjects and 20 additional (repeat) measurements. A mixed models approach with fixed and random effects is used for analysis [SAS PROC MIXED]. From this model, a coefficient of variation between technicians can be estimated, the possibility that a single examiner differs from all other examiners can be tested, and the relative contribution of examiner and participant variation to the overall variation can be calculated.

Depending on whether the exam is fatiguing, i.e., the first measurement likely to affect the value of the second measurement, the exam is either repeated on another day soon after the first exam, or on the same day. In either case, a different examiner is used for the second measurement.